Statistical Analysis via the Cauchy Distribution: a New Method with an Application to Electoral Data

Ivan H. Krykun[1,2]

[1] Department of Theory of Control Systems, Institute of Applied Mathematics and Mechanics NAS of Ukraine, Sloviansk, Ukraine

[2] Department of Information Management, Mathematics, and Statistics, «KROK» University, Kyiv, Ukraine

Email address:

iwanko@i.ua

Abstract

Many natural phenomena are frequently influenced by multiple random factors which can often be plausibly modeled as normally distributed random variables. This paper proposes a novel method for the statistical analysis of such phenomena, exemplified through electoral processes. Election outcomes are determined not only by the stochastic support for various candidates but also by voter turnout, which can occasionally be a decisive factor. Furthermore, this study yields a tool for estimating election fraud.

Keywords

Arctangent regression, Cauchy distribution, electoral statistics, statistical estimation of parameters.

I. Introduction

Many natural phenomena depend substantially on random factors that can be modeled as normally distributed random variables. Using electoral processes as an example – where outcomes are determined by both stochastic candidate support and decisive voter turnout – this paper proposes a new method for the statistical analysis of such phenomena.

Elections are an important and interesting example of the practical implementation of the random behavior of human society. After all, the results of elections depend not only on, generally speaking, random and variable support for candidates, but also to a certain extent on the – also random – turnout of voters, which, in the case of approximately equal support for candidates, becomes decisive.

Several approaches have been proposed by researchers to detect possible electoral fraud.

First, the authors look at the frequency distribution of turnout or candidate's results across electoral districts. If the data are genuine, we would typically expect this distribution to be normal. Deviations from normality – such as bimodality or a substantial tail of districts with very high turnouts – give cause for suspicion. Also on such frequency distribution of turnout is noticeable increasing of round numbers (such as, e.g., multiples of 5 or 10).

Second, the authors look at how each candidate's share of the total electorate correlates with the turnout. If voting figures are reported accurately, we would expect a candidate's share of the electorate to increase with turnout in rough proportion to his or her overall vote share: if turnout increases by 100 votes, we would expect a candidate who scores 50% of the vote overall to pick up around 50 of these votes. If, by contrast, turnout is inflated by the casting of ballot boxes, the proportion of additional votes captured by the favoured candidate will be far higher.

The third approach is the analysis of the ratio of votes cast by candidates/parties, depending on the turnout. It was first used by J. B. Kiesling to analyse elections in Armenia. The method has proven itself useful in Russia to identify fraud in favour of the ruling party and its candidates.

Another approach is the use of Benford's law (or the first-digit law) to compare the obtained frequency of appearance of the first (or second) digit in the final election documentation with the theoretical distribution.

A more detailed review of statistical approaches can be found in [1].

II. Main Result

The new method of statistical analysis of some natural phenomena proposed in this work is based on the well-known result that the ratio of two standard normal (i.e. with parameters 0 and 1) random variables is a random variable that follows the Cauchy distribution with parameters 0 and 1.

We recall that random variable is said to have a Cauchy distribution with parameters α and $\gamma > 0$ if its cumulative distribution function is

$$F(x) = \frac{1}{2} + \frac{1}{\pi} arctan\left(\frac{x-\alpha}{\gamma}\right),$$

and its probability density function is, respectively,

$$f(x) = \frac{1}{\pi}\frac{\gamma}{(x-\alpha)^2+\gamma^2}.$$

It is well known that the Cauchy distribution does not have moments of any order and belongs to the class of heavy-tailed distributions for which the law of large numbers does not hold. Due to these properties of the Cauchy distribution, it is not possible to estimate its parameters by standard methods of mathematical statistics (such as method of moments or maximum likelihood estimation), several different methods have been proposed for this.

To address this, an alternative approach for estimating the parameters of the Cauchy distribution was developed by the author in previous work [2]. The core idea involves constructing an arctangent regression after specific processing of empirical data and then using the least squares method to estimate the parameters of this regression, which correspond to the parameters of the Cauchy distribution. Based on this approach, the following proposition for a new method of statistical analysis is formulated as follows:

Proposition. [3, Proposition 1] After centering and normalizing selected indicators of natural phenomena (which are assumed to be normally distributed), we construct the ratio of two such indicators. The resulting data is then statistically processed according to the procedure described in [2] to obtain empirical estimates of the distribution parameters. These estimates are then compared with the theoretical values of the Cauchy distribution parameters (i.e., 0 and 1). In case of significant discrepancies, the probability of such a deviation is assessed, and statistical conclusions about the analyzed phenomena are drawn.

III. Example and Conclusion

This proposed method was applied to the statistical analysis of the election results from the first round of the 2004 Presidential Election of Ukraine in the districts of the Donetsk and Luhansk regions in [3]. The analysis revealed statistically significant deviations in the estimated parameters from the theoretical values, suggesting the presence of irregularities in the electoral data.

This approach demonstrates potential for robust anomaly detection not only in electoral analysis but also in other fields where processes are driven by multiple normally distributed random factors. The proposed anomaly detection method also shows significant promise for industrial applications. It could be instrumental in quality control for manufacturing processes where the ratio of two normally distributed measurements is a key performance indicator, including pharmaceutical dosage, chemical compound synthesis, and metallurgical alloy composition.

Funding

This work was partially supported by a grant from the Simons Foundation (SFI-PD-Ukraine-00017674, Krykun I.H.)

Acknowledgment

The author expresses his sincere gratitude to the heroic soldiers of the Armed Forces of Ukraine, who have been protecting the author, his family, and his friends from the bloody Russian terrorist forces since 2014.

References

1. M. Myagkov, The Forensics of Election Fraud: Russia and Ukraine. New York: Cambridge University Press, 2009.
2. I. H. Krykun, The Arctangent Regression and the Estimation of Parameters of the Cauchy Distribution, Journal of Mathematical Sciences 2020, 249, Issue 5, pp. 739–753.
3. I. H. Krykun, New Approach to Statistical Analysis of Election Results, International Journal of Mathematical, Engineering, Biological and Applied Computing, 2022, Vol. 1 № 2. pp. 68–76.